Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

INFORMATION VISUALIZATION SUPPORTING MODELLING AND EVALUATION TASKS FOR CLIMATE MODELS

Thomas Nocke
Heidrun Schumann

Institute of Computer Graphics
University of Rostock
D-18051 Rostock, GERMANY

Uwe Böhm

Institute for Physics
University Potsdam
Am Neuen Palais 10
D-14469 Potsdam, GERMANY

Michael Flechsig

Potsdam Institute for
Climate Impact Research
P.O Box 60 12 03
D-14412 Potsdam, GERMANY

ABSTRACT

Information visualization exploits the phenomenal abilities of human perception to identify structures by presenting abstract data visually, allowing an intuitive exploration of data to get insight, to draw conclusions and to interact directly with the data. The specification, analysis and evaluation of complex models and simulated model data can benefit from information visualization techniques by obtaining visual support for different tasks. This paper presents an approach that combines modelling and visualization functionality to support the modelling process. Based on this general approach, we have developed and implemented a framework that allows to combine a variety of models with statistical and analytical operators as well as with visualization methods. We present several examples in the context of climate modelling.

1 INTRODUCTION

Fast developments occurred in the fields of data acquisition and modeling of complex systems. On the one hand, many new measurement techniques with increasing resolutions generate large amounts of data. On the other hand, increasing computing power of massive parallel supercomputers generate large simulated data sets. When trying to empirically identify the underlying properties of these large data sets, such as patterns and relations between the variables, exploration and diagnosis become more and more the bottleneck. Our approach is to apply new methodologies from information visualization as support to handle, analyze and evaluate these large data sets effectively. This will be outlined on the example of climate modeling.

Our paper focuses on the combination of visualization methods with analytical mathematical and statistical techniques for model specification, analysis and evaluation. Such strong linkage of these techniques is still an open problem. However, the exploration of real-world data and model results may strongly benefit from visual support. The combination of automated and visual methods is a current research topic called visual data mining and allows to extract hidden knowledge from the data intuitively (Keim, Müller, and Schumann 2002).

Although visual data mining primarily focuses on the exploration of a data set, it can also be applied to different tasks of the modeling process. For example for model specification purposes, combined visualization and statistical methods can be applied to acquire a-priori knowledge from raw data, and to formulate hypotheses e.g. on regularities or trends. This may help to qualitatively detect these patterns within the data, to pre-select suitable statistical methods, to quantify them and to specify a suited model.

In the context of climate modeling, this combination may provide substantial support in identifying reduced climate models. Such reduced climate models (Pedlosky 1987) can be used to study specific processes, which are dominating the behavior of the climate system under specific conditions. Acquiring and exploring metadata from the model results and applying visualization methods to these results may help to identify essential system variables for the process of interest, and therefore to reduce the degrees of freedom in a complex physical model.

Furthermore, appropriate visualization techniques are also able to improve the analysis process for a given model. They can be applied to check, if the underlying mathematical model is properly transformed into a numerical one by discretizing the model equations and applying suitable solvers.

Finally, expressive visualization methods can be utilized for model evaluation. Here, an a-priori knowledge of sensitive sub-processes and model variables resulting from the application of data mining tools can be helpful to pre-select suitable statistical methods for quantifying model errors.

A combination of methods to support the described tasks may remarkably reduce the required analysis efforts and strongly increase the efficiency of any data analysis.

Solving these tasks requires special care. In this context, a visualization tool has to consider the following aspects:

- a high degree of interaction and navigation functionality to handle the model output,
- capabilities to handle high dimensional and heterogeneous data input (including state space, parameter space, physical space and time reference),
- a high degree of user support, and
- a general applicability and reusability for different kinds of scenarios, model types and scientist's intentions (in climate research).

Visualization tools, which provide the functionality listed above, can really support model specification, analysis and evaluation. The paper will discuss these aspects in more detail, and is organized as follows: First, we give a brief overview of the related work (section 2). After a short problem discussion we will give a general outline of our approach (section 3). Section 4 demonstrates visual support for climate modeling by several examples. Finally, we will conclude our contribution with some remarks on further work in section 5.

2 RELATED WORK

In the last years, new challenges have been encountered for information visualization, ranging from the strongly increasing computing power that nowadays allows to specify more complex models and to run existing numerical models with much finer discretization steps. To visualize the huge amount of data resulting from such simulations requires to reduce the complexity of data by post-processing them and applying new visualization approaches. Many techniques are already available to generate e.g. map representations of results and errors for presentation purposes. However, widespread software packages as GrADS or Vis5D (Doty and Kinter 1995; Hibbard et al. 2000) often require specific features with respect to the data grids and are adapted to special purposes and map representations. Also, there is a lack of user-friendly interaction possibilities to extract information from data and to identify sensitive processes and their temporal and spatial dimensions. Furthermore, the problem of pre-selecting suitable techniques from these frameworks according to the characteristics within the data is still not satisfyingly solved.

However, information visualization provides many useful methods to solve the mentioned tasks. Combined with well-known mining techniques such as association rules, self organizing maps and clustering techniques (Han and Kamber 2001; Keim, Müller, and Schumann 2002) it really can support the modeling process. A variety of approaches and systems have been developed, basing on the tight linkage of information visualization and non-visual mining techniques (Brunk, Kelly, and Kohavi 1997; Westphal and Blaxton 1998; Kreuseler, Nocke, and Schumann 2003). Nevertheless, a universal support and linkage of visualization and non-visual mining methods for complex system modeling tasks is still an open research topic.

Moreover, comparing data, for instance real-world data with simulated data for evaluation purposes, often is done by linking & brushing (Unwin, Wills, and Haslett 1990). Besides this, there exist some special approaches. For instance, Pagendarm and Post (1997) introduce an approach of comparative visualization of flow data sets on varying grids. However, general approaches for the visual comparison of data and models are not available and have to be developed.

In our paper we will concentrate on climate research as a specific application field. Especially for the evaluation of climate models, a wide range of measures exist that were successfully applied for individual variables during the past. For instance, these measures can be used to quantify model errors (e.g. Jones, Murphy, and Noguer (1995)). Furthermore, multivariate statistical methods have recently been combined with multidimensional error measures (Böhm 1999; Kücken, Gerstengarbe, and Werner 2002) which can be utilized in our research to comprehensively investigate variable combinations of real world data or of model data representing the major players in complex processes. Thus, they can provide the required input for visualization techniques to investigate the usefulness of visual support.

3 PROBLEM DISCUSSION AND GENERAL APPROACH

On the one hand, there exist a variety of methods for the specification, reduction, evaluation and comparison of models in different domains. These methods are more or less integrated in general tools. On the other hand, there are many systems and frameworks for information visualization. The problem in this context is, that a simple linkage of these partly complex systems is not possible, and a separate processing is time consuming, not intuitive, and restricted to low-level task communication. Thus, a flexible, easy-to-use framework is needed, tightly linking visualization and modeling methods.

There are 3 possible ways to achieve this goal:

- The functionality of an appropriate visualization system has to be extended to consider the specifics of the modeling task.
- The functionality of a modeling tool of interest has to be extended with advanced visualization techniques to extend or replace the more simple graphical output facilities of today's modeling systems.
- A new framework has to be designed, which integrates both, suitable modeling and visualization tools.

Usually, general modeling systems as well as information systems provide an expanded functionality, which seldom is necessary in its entirety for a given task. Advancing such systems with more complex additional facilities leads to very comprehensive, and expensive systems. Therefore, a compact framework with well defined modeling and visualization functionality can be the better way. However, in this case special purpose-solutions are developed, and it is difficult to adapt them to other problems.

Our work concentrated on two of the three strategies. First we used the visualization system OpenDX (IBM) as software platform, and enhanced it with additional functionality for our purposes. We chose OpenDX, because it is an efficient platform-independent tool. It provides nearly the same functionality as other data-flow oriented visualization systems such as AVS. It is public-domain, a variety of default techniques are available, and in general, new techniques can be integrated easily. In contrast to other systems like MatLab more visualization functionality as well as interactivity is provided. However, there are still limits to the application of OpenDX:

- *Graphical user interface*: Unfortunately, the number of GUI-elements in OpenDX is limited (to simple elements), leading to the following problem: in complex visualization networks the number of interactors and the number of their option dialogs is increasing. This results in dialog overlapping and user confusion for parallel exploration tasks. Structural GUI-elements such as tabbed panes are missing to handle this problem.
- *Temporal efficiency*: OpenDX' internal processing mechanisms are limited in their run-time performance in processing large data sets (e.g. because of data structures, caching mechanisms, looping processing and image rendering). Often, an implementation of special visualization models for instance in OpenGL could be a better solution to handle large data sets effectively, directly controlling data flow and output of graphic primitives.
- *Interaction behaviour*: Furthermore, default OpenDX' piking and interaction functionality is limited to simple point wise clicking. Complex interactions like "on mouse move" events are not supported in current version 4.2.0.

Nevertheless, the application of OpenDX is an appropriate way to support the modeling process by visualization.

As a second strategy we designed a new framework, which integrates modeling and visualization abilities. Therefore, conceptually all the available functionality of OpenDX can be capsulated. However, in general we integrated compact visualization techniques suited for modeling tasks. To ensure a high flexibility, visualization techniques developed with other systems (e.g. OpenGL) can be capsulated as well. A more detailed discussion of our framework will be presented in section 4.4.

Hence, we want to suggest a general approach for the design of such frameworks, combining modeling and visualization tools. First we have to specify the general functionality of such a system. For this purpose, we have identified 3 different tasks, which have to be supported by visualization techniques:

- Specification of a model
  Here, real world data have to be visualized to derive first hypotheses. Furthermore, visual support has to be given for the derivation and adjustment of model parameters, or equations for essential system variables.
- Analysis of a model
  This includes the visual exploration of the input and output data of a model, as well as the visual exploration of the model itself to support the identification of internal processes.
- Model evaluation
  Here, visual representations are used to show the features of a model for evaluation purposes. For example the joint visualization of real world data with simulated data allows a visual comparison of the characteristics of both data sets, and in doing so, an evaluation of the model.

It is obvious that visualization methods supporting these tasks have to possess a high degree of interactivity. Furthermore, it is obvious that the problem-specific domain has to be given to specify the necessary visualization and modeling functionality for these tasks.

In our case, the background is the specification, reduction and evaluation of climate models. Therefore, the following functionality regarding climate modeling has to be supported by visual means:

- Visual exploration of real-world climate data and of complex models to identify internal processes and features: Since the given climate data are very comprehensive, e.g. clustering is used, and thus, cluster methods have to be combined with advanced visualization techniques.
- Derivation, modification and effective calculation of reduced climate models: for this task visual support can also be given. This includes the possibility of the interactive definition of discretized ordinary differential equation systems (ODES). For this purpose the development of an interactive tool is necessary, that intuitively encapsulates parser, solver and error estimation functionality, allowing to calculate ODES with any number of equations and visually representing the output. Furthermore, this includes the visual specification of input parameters such as starting values.
- Comparison and evaluation of a complex and a reduced model: Here, methods for a joint visualization of the parameters and data values of the original and the reduced model as well as the joint visualization of model data and real-world data have to be applied.

In general, information visualization focuses on the representation of an abstract data set. In the context of supporting modeling processes there are some additional requirements, which have to be considered. In the following we present visualization methods for climate data, and discuss how these methods can help in the tasks listed above. Although these methods are designed especially for climate modeling, they can demonstrate how useful visual support can be in modeling processes in general.

4 INFORMATION VISUALIZATION FOR CLIMATE MODELLING

In this section, we provide examples to demonstrate the feasibility of using advanced visualization algorithms for modeling and quality assurance of model results. Furthermore, an overall framework for all these techniques is presented that allows users to easily combine individual methods in an interactive way.

4.1 Model Specification

This task includes the visual exploration of real world data to support the derivation of models from the gathered knowledge. Therefore, in this stage a variety of interactive information visualization techniques have to be provided together with appropriate preprocessing techniques to handle huge amounts of data. Statistical methods such as clustering can be applied to gain several abstraction levels of the data. This allows the interactive exploration of large data sets, starting with a compact overview image - avoiding data overlapping - and then adding more details interactively. Furthermore, to handle large data sets in this context, selection mechanisms from information visualization such as brushing & linking can be applied.

Figure 1 gives an idea of this approach (based on the calendar cluster view of van Wijk and van Selow (1999)). Here, clustering of daily temperature data was provided to reduce the amount of the underlying data set for a first exploration step. For the graphical representation of the clustered data, two different views are used and linked by brushing.

The first view is a calendar plot (left), color-coding the cluster identifiers of each day with a certain color. The second view is a trend plot (right), representing the temperature functions of the different clusters, days and months. Several interaction functions are included to support the exploration process. For example, this includes the selection of a mediod (mean trend) or a median (typical day representing its cluster) modus. Furthermore, by selecting and deselecting clusters, days and months, special information are shown or hidden. This allows an effective evaluation of certain clusters, days and months as well as of their (mean) daily cycles and their extremes, avoiding overlapping. Thus, the user can decide by himself if the image is still effectively perceptible.

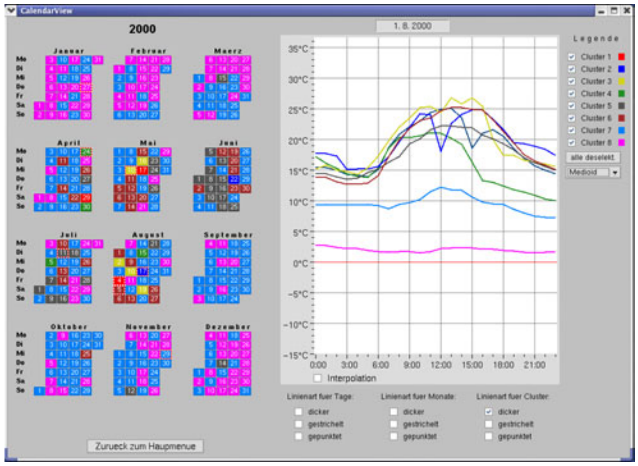

Figure 1: Cluster Visualization of Daily Temperature Cycles based on Hourly Measured Observations at the Potsdam Station during the Year 2000, using Correlation Coefficient as Similarity Measure and the Cluster Calendar View of van Wijk and van Selow (1999)

In doing so, typical and non-typical clusters and relations can be identified. This helps to better understand the current climate and to improve existing models. In this example, the clusters 7 and 8 represent typical daily cycles. All the other clusters are more or less “non-typical”, so-called outliers. Furthermore, the color-coded calendar allows to reveal fast changes in cluster sequences for example in the first part of August (see figure 1). Altogether, this technique allows the comparison of cluster cycles (overview), the exploration of single cluster cycles (abstract detail) and the exploration of daily and monthly values of interest (specific details). Finally, this approach allows to link specific daily cycles of interest to the whole data set to recognize significant changes over time. Such a merging of empirical results with background knowledge can be seen as one step towards an empirical model formulation.

We want to give another example to show the effectiveness of visual exploration. To analyze extreme hot summers at the observation station Potsdam, the years from 1893 to 1997 have been characterized by 5 parameters, describing extreme summer conditions. Afterwards, the years have been clustered and visualized (see figure 2).

Figure 2a shows all clusters, using a default arrangement of 10 years per column. At the first glance, we can see the increasing number of years belonging to cluster 2 in the second half of the 20th century, representing the most extremely hot summers.

By interactive modification of the arrangement of the clusters we can get deeper insight into the data. For example, placing 6 years per row indicates a certain periodicity of cluster 2. While a high accumulation of years in which cluster 2 appeared in the first to third column can be observed, the frequency of this cluster is very low in the other columns (0, 4, 5). The so-identified quasi-6-year cycle be-

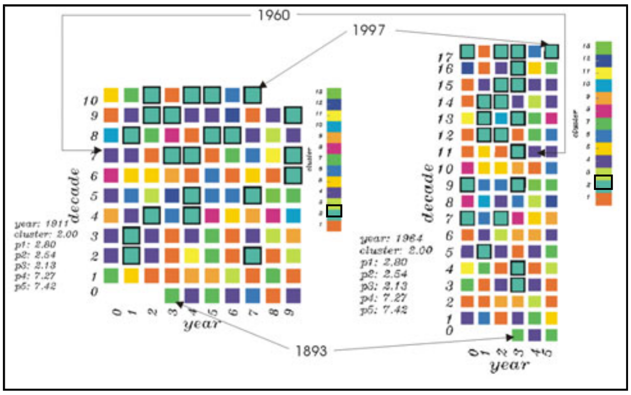

Figure 2: Temporal Classification of Time Series for the Meteorological Observation Station Potsdam (OpenDX); Parameters are Derived from Daily Values for Temperature; a) Period of 10 Years (left);
b) Period of 6 Years (right)

comes visible during the early sixties and seems to disappear after 1990, when extremely hot summers occur more often. This implication can be statistically proven as well (e.g. with an autocorrelation function) and gives an indication on the way how the transition from a stable climatic state to a new one proceeds at this station, also leading to empirically derived model assumptions for this process.

In a way as shown in the two examples above, new models can be specified and modified. The next example illustrates how visualization may support the modification of an already existing model. Here we developed an interactive tool for the specification and parameter optimization of a discretized ODES (see figure 3).

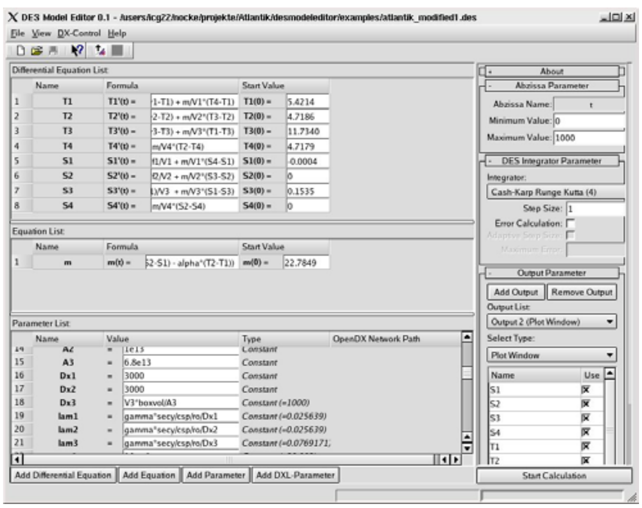

Figure 3: DES Model Editor: Support for the Interactive Definition of ODES-Based Discrete Models (Example: a Simplified Atlantic Model Consisting of 8 Variables (4 Temperature and 4 Salinity Variables Representing the Mean Values of 4 Atlantic Regions))

In dependency of the visual exploration process, an arbitrary number of discretized differential equations (top), initial and boundary conditions (center) as well as time dependent auxiliary conditions and constant parameters (bottom) can be defined, calculated and displayed. Furthermore, the interface supports the specification of model and experiment dependent metadata such as step size and number of integrations (top right). Different solvers have been integrated in our tool, including error calculation and adaptive increment (right center). Of special interest for the modeling process is the strong connection with visualization capabilities.

On the one hand, visual representations are used to specify and optimize parameters for the discretized model equations. Figure 4 shows an example for this. Each of the four boxes represents a region of similar properties for the driving variables forcing the Atlantic oceanic circulation. These boxes can be visually defined and adapted. The proper choice of the box areas is essential for obtaining realistic results because certain initial conditions and other model parameters depend on these terms.

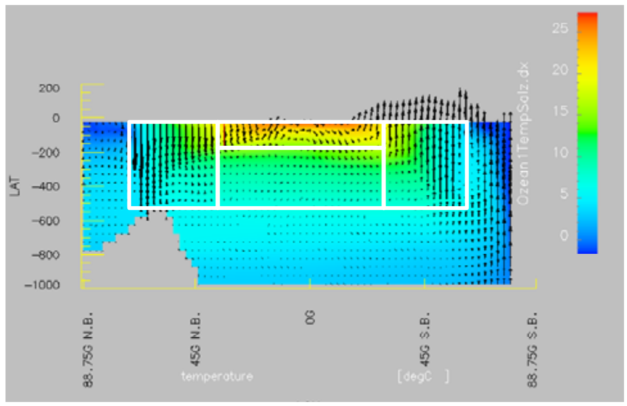

Figure 4: Exploration of Model Results and Specification of Parameters for a Simplified Model (OpenDX): Area Definition in the Atlantic Ocean Part (Vertical Depth-Latitude Cut) of the Climate Model CLIMBER-II (Petoukhov et al. 2000) based on the Temperature Gradient Field

On the other hand, a suitable visualization of model output data supports the model analysis and evaluation. This includes the interactive parameterization as well as the selection of sets of variables and of error measures (see figure 6 in section "Model evaluation"). Thus, the selected variables and their temporal and spatial behavior can be visually explored and evaluated.

4.2 Model Analysis

After the model specification process is passed, the behavior of a model, for instance in terms of its stability to disturbances, has to be analyzed. In climate models, in many cases one has to cope with 3d-temporal time series of simulated data with more than 100 parameters. Handling this amount of data can strongly benefit from visualization. In the following, two examples for model analysis and their advantages in this context will be presented.

The first example addresses the exploration of the Atlantic temperature conditions over a period of 700 years (see figure 5). In a volume-rendered image, temporal temperature changes (in relation to the starting values) are color-coded. The initial temperature values in the depth-latitude plane are mapped to a grayscale color map, supporting the comparison of absolute temperature values and their changes. Positive changes are mapped to a yellow-red color map, and negative changes to a blue-cyan color map.

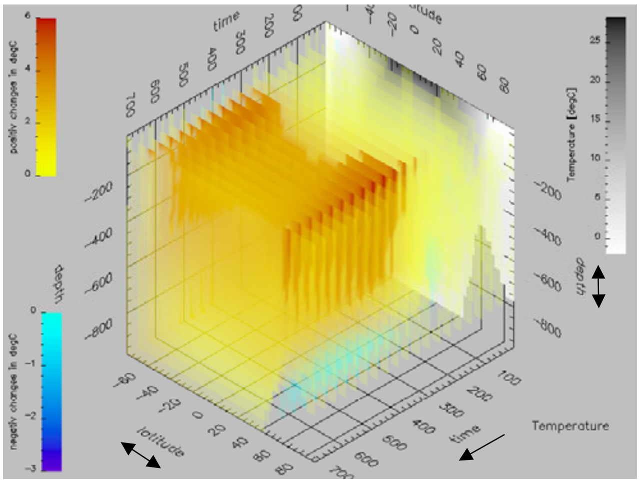

Figure 5: Volume Rendered Image (OpenDX) of Temporal Temperature Differences of the Atlantic Part of the CLIMBER-II Model (Petoukhov, Ganopolski, Brovkin, Claussen, Eliseev, Kubatzki, and Rahmstorf 2000)

In detail, figure 5 displays a depth-latitude cut (x,y-axes) in the Atlantic ocean over the time (z-axis). On the one hand, yellow-red areas in the diagram represent regions of temperature increase in relation to the initial state. Especially in the time period from year 400 to year 700 in the upper ocean regions a strong temperature increase (more than 3 Centigrade in dark red) can be identified. On the other hand, cyan areas in the deep regions of the northern Atlantic exhibit a slight temperature decrease with time.

In addition, to focus on certain variable ranges (such as a strong temperature increase in fig.5), the opacity of the displayed changes can be adapted interactively. Thus, variable ranges of interest can be focused. For instance in figure 5, the opacity function in the attribute range between 0.0 and 3.0 degrees centigrade is increased. Thus, hidden regions of highly positive changes and their range become clearly visible.

Altogether, this technique provides a good visual support for different user goals by hiding of some and focusing on certain other variables and variable changes in combination with variations of the viewpoint. The challenge in this context is the automated parameterization of color maps and viewpoints in dependency of user goals and data characteristics.

A second example for model analysis is the comparison of two different simulations of the same model under modified initial conditions.

Figure 6 gives an idea of this approach. It shows the output of a simplified model for the Atlantic oceanic circulation, represented by four boxes covering regions with similar properties of the driving variables (see fig. 4).

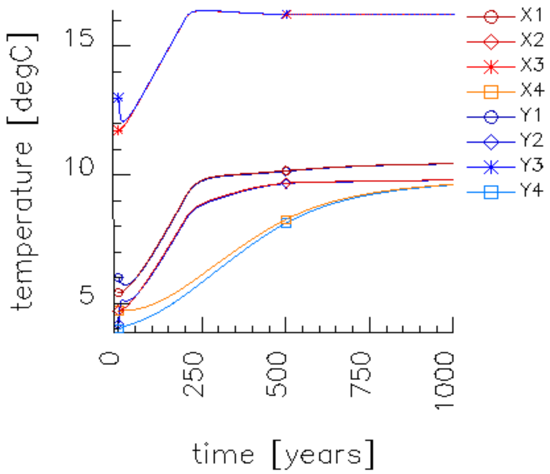

Figure 6: Temporal Evaluation of Two Different Model Runs (Run X: Red, Run Y: Blue) of a Simplified Atlantic Model with 4 Temperature Variables

The red bundle of graphs represents the results within the four boxes for the reference run (X), while the blue trajectories represent a run with interactive adaptation of the starting parameters (run Y). The results for the individual boxes (1 to 4) can be separately identified, applying function markers and differing brightness. Thus, stability and convergence can be evaluated. For instance temperature values within the boxes 1, 2 and 3 (X1 and Y1, …) converge for both runs in less than 100 integration steps. In contrast, the variables X4 and Y4 become similar only after more than 500 hundreds integration steps.

In addition, this technique supports an interactive selection mechanism to get a deeper insight into interesting details. This allows to reveal and to hide several individual graphs, avoiding image overloading. For instance, this helps to distinguish the specific curves in regions of high graph density more clearly (see. fig. 6, e.g. the area between 0 and 5 degrees centigrade in the first 100 time steps).

4.3 Model Evaluation

Model evaluation is of utmost importance to assess a model's performance qualitatively and quantitatively. This includes for example the comparison of simulated and measured data as well as the inter-comparison of models (e.g. the evaluation of a simplified model with a complex climate model). Challenges in this context are the amount of data and the comparison of data on different grids.

An often-applied example for model evaluation is the comparison of calculated and measured data sets. Here, we assess the potential total yield loss for maize in the North-east of Brazil due to water scarcity. Agricultural experts provided estimates for idealized precipitation thresholds for the various phases of the growing season, which are required for a minimum yield of maize crops. These criteria have been used to scale both observed rainfall data and model results that were interpolated to station sites. We have performed a cluster analysis (Gerstengarbe and Werner 1999), separately for both model and reference data, using the resulting parameters. The identified clusters have been set in relation to each other by Euclidean distance measures and the most similar clusters from both data sets were visualized using the same colors. Figure 7 compares the clustering results for the station observations and the model results in spatial Voronoi diagrams.

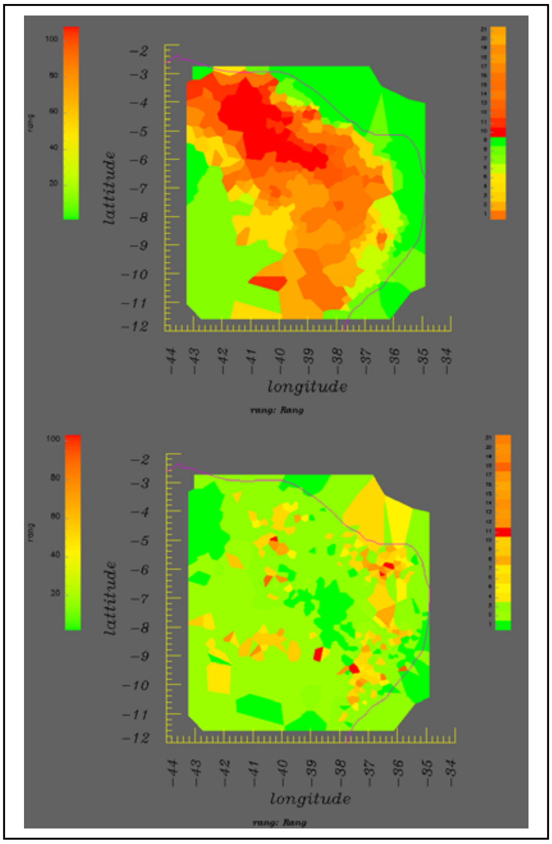

Figure 7: Visualization of the Clusters Expressing the Potential Total Yield Loss for Maize in the Northeast of Brazil in 1983 in their Spatial Reference (Voronoi Tessellated Areas; Color Coding of Aggregated Variable Rang; OpenDX);

a) Clustered Station Observations (top);
b) Clustered Climate Model Results, mapped to Measurement Station Locations (bottom)

This technique can be interactively re-parameterized with other color maps, displaying other yield loss parameters or the color-coded cluster identifiers to improve the knowledge of spatial cluster distribution and of cluster properties. Fading in orography information and measurement locations provide additional information in this context. Especially in this example, the areas of the most pronounced drought patterns and their intensity vary remarkably, whereas in coastal and less vulnerable regions similar results become visible. Based on this kind of exploration, we could trace back the sources for these model deficiencies and identify an inadequate representation of hydrological processes in the soil. Furthermore, model intercomparison becomes possible using the same distance measures to the reference data for various other models.

4.4 The Framework

As described in section 3, there are still some limitations applying common visualization systems such as OpenDX for modeling and evaluation tasks. Thus, a new framework has been designed and implemented. This framework integrates a variety of modeling, statistical, analytical and visualization methods including the functionality described above.

The framework supports data input from several file formats (OpenDX, NetCDF, ASCII). Metadata from these file formats are imported and stored in an internal data format based on our metadata specification (Nocke and Schumann 2002). Internal data flow and operator states (e.g. the information, if an operator can be executed or not) is based on metadata of the data inputs at the certain operators. For instance a 3D-Visualization operator will be executed only, if data input contains valid 3D data.

The visualization functionality of the OpenDX has been wrapped to visualization operators (using the DX-Link library). This includes wrapping of standard techniques such as isosurfaces as well as wrapping of extended and new visualization techniques in OpenDX. Furthermore, visualization techniques based for instance on OpenGL or QT (such as the ODES editor), and data mining techniques such as clustering can be wrapped as well.

To integrate the described functionality in an easy-to-use environment, we use the QT-Library from Trolltech (Trolltech 2003). We keep the data flow paradigm of OpenDX, but simplifying it by using compact visualization operators instead of the low level modules of OpenDX. In doing so, we avoid overloading the user interface with visualization specific and internal aspects. This approach allows a tight linkage of techniques at a high abstraction level, avoiding low-level user interception such as data conversion.

Beside handling of different windows and menus and their communication, the application of the QT library extends the possibilities of GUI interaction paradigms of OpenDX. This includes the design and implementation of a dynamic parameter dialog, encapsulating the set of parameters in one window.

Figure 8 displays a screenshot of the framework main window. Here, the general functionality is provided, including visualization, statistical, and analytical operators. They are organized in directed graphs, where the edges represent data flow between inputs and outputs of certain operators (right window). Problem specific adaptations of well-known layout algorithms have been integrated to support graph and edge layout.

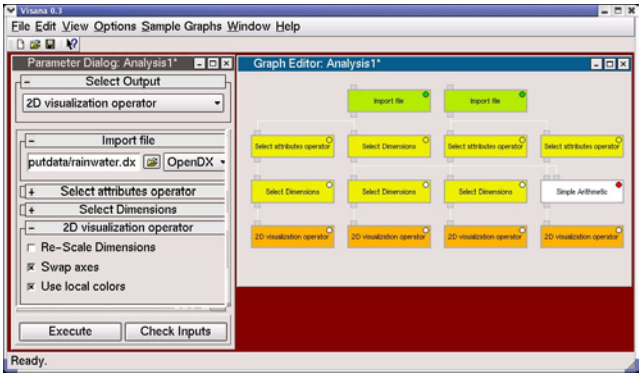

Figure 8: A Framework Screenshot: Dynamic Parameter Window (left); Graph Editor (right)

The user can interactively generate, replace and delete operators and data connections. Furthermore, direct feedback about operator states (colored circle in the top right corner of each operator) and operator types (operator color) is given.

The left part of figure 8 shows the operator parameter dialog of the current graph. The design goal of this dialog is to handle many operator parameters even for large graphs, by integrating all the parameter windows into a single one. Because of the complexity of the dialog and the high amount of operator parameters, the dialog window is organized as follows: Applying the topmost selector, all the parameters for a certain visualization or other output operator can be revealed. Furthermore, parameters can be hidden and revealed for each operator. The advantage of such a dialog is that it avoids user confusion and window overlapping, if there are more than one visualization window and several parameter windows active.

To support the reuse of graphs for similar tasks and the storage of gathered knowledge, graphs and their parameter sets can be stored, reloaded, evaluated and commented. Altogether, a flexible and easy-to-use interface based on a flexible, object-oriented software design in combination with powerful metadata and operator concepts enables the users to focus on the modeling, analysis and evaluation tasks only.

However, the framework is still under development. We tested it for different problem cases. For instance we tested the ODES-Operator with an simplified Atlantic circulation model with 8 differential equations (4 describing salinity and 4 describing temperature conditions in certain regions of the circulation), one non-differential equation (describing the mass flow) and 68 parameter functions and constants. Using the Fehlberg-Runge-Kutta integration technique for 100.000 time steps took about 5.7 seconds on a 800MHz-Pentium processor. This amount of time is still acceptable for interactive modeling of ODES.

5 CONCLUSION AND FUTURE WORK

In this paper, we investigated the possibilities of interactive coupling of modeling functionality with information visualization techniques. Several visualization techniques supporting model specification, model analysis and model evaluation for climate models have been presented. A general approach for a framework and its implementation, realizing the described functionality, has been outlined.

Nevertheless, there are still challenges for future work: On the one hand, further visualization techniques have to be developed to increase the visual support for real-world data exploration as well as for model analysis and for model comparison of complex models. This includes the visualization of very large, heterogeneous data sets, the visualization of model comparison on different grids with a high number of parameters and the visualization of errors and uncertainties in space and time.

On the other hand, these new techniques have to be coupled with modeling methods and integrated into our framework to enhance its functionality. Although first experiments with our framework with simple models were very successful, and we got first new ideas about the data, we have to continue testing and evaluating its usability.

ACKNOWLEDGMENTS

The authors thank Henrik Kinnemann for his motivation and ideas, which established an interesting cooperation of the PIK and the ICG Rostock. Furthermore, we thank Rupert Klein for his impulses and proposals to this interdisciplinary paper, and Georg Fuchs for proofreading our English. Finally, we thank all the students working on the framework for their dedicated involvement. These are Mathias Holst, Martin Schmidt, Torsten Struck, Hagen Marczok, Karsten Wagenknecht, Matthias Schröder, Eik Deistung and Marc Alinski.

REFERENCES

Böhm, U. 1999 (in german). Eine Methode zur Validierung von Klimamodellen für die Klimawirkungsforschung hinsichtlich der Wiedergabe extremer Ereignisse; PhD thesis, Freie Universität Berlin, Fachbereich Geowissenschaften

Brunk, C., Kelly J. and Kohavi, R. 1997. MineSet : An Integrated System for Data Mining, *Proceedings of the* *third international conference on Knowledge Discovery and Data Mining*

Doty, B.E. and J.L. Kinter III. 1995. Geophysical Data Analysis and Visualization using GrADS. *Visualization Techniques in Space and Atmospheric Sciences*, eds. E.P. Szuszczewicz and J.H. Bredekamp, NASA, Washington, D.C., 209-219.

Gerstengarbe, F.-W. and Werner, P.C. 1999. The complete non-hierarchical Cluster-Analysis; *PIK-Report* No. 50, Potsdam.

Han, J. and Kamber, M. 2000. *Data Mining: Concepts and Techniques*, 8 ed., The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers.

Hibbard, B. et al. 2000. Vis5D Documentation [online]. Available online via <http://vis5d.sourceforge.net/vis5d.pdf> [accessed Juli 1, 2003].

Jones, R. G., Murphy, J. M. and Noguer, M.1995. Simulation of Climate Change over Europe using a nested regional climate model; I: Assessment of control climate, including sensitivity to location of lateral boundaries. Q. J. R. Meteorol. Soc., 121, pp. 1413-1449.

Keim, D., Müller, W. and Schumann, H. 2002. Information Visualization and Visual Data Mining; State of the art report, *Eurographics 2002*, Saarbrücken, Sept.

Kreuseler, M., Nocke, T. and Schumann, H. 2003. Integration of Clustering and Visualization Techniques for Visual Data Analysis.; In: Opitz, O.; Schwaiger, M., (ed.), *Exploratory Data Analysis in Empirical Research, Studies in Classification, Data Analysis and Knowledge Organization*, Springer-Verlag, Berlin-Heidelberg.

Kücken, M., Gerstengarbe, F.-W. and Werner, P. C. 2002. Cluster analysis results of regional climate model simulations in the PIDCAP period. *Boreal Environment Research* Vol.7, No 3.

Machenhauer, B., Windelband, M., Botzet, M., Christensen, J. H., Déqué, M., Jones, R. G., Ruti, P., M. and Visconti, G. 1998. Validation and Analysis of Regional Present-Day Climate and Climate Change Simulations over Europe; MPI-Report No. 275. Hamburg.

Nocke, T. and Schumann, H. 2002. Meta Data for Visual Data Mining ; Proceedings Computer Graphics and Imaging, *CGIM'02*, Kaua'i, Hawaii, USA, August, 2002.

Pagendarm, H.-G. and Post, F. H. 1997. Studies in Comparative Visualization of Flow Features, In: G. Nielson, H. Hagen and H. Mueller: *Scientific Visualisation. IEEE Computer Society*, Los Alamitos, pp. 211-228.

Pedlosky, J. 1987. Geophysical Fluid Dynamics; 2nd ed., Springer, Berlin.

Petoukhov, V. A., Ganopolski, A., Brovkin, V., Claussen, M., Eliseev, A., Kubatzki C. and Rahmstorf, S. 2000. CLIMBER-2: A climate system model of intermediate complexity. Part I: Model description and performance for present climate. Climate Dynamics 16, pp. 1-17.

Rahmstorf, S. 1996. On the freshwater forcing and transport of the Atlantic thermohalince circulation. *Climate Dynamics*, 12, pp. 799-811.

Trolltech. 2003. Qt 3.1. Whitepaper [online], Available online via: ftp://ftp.trolltech.com/qt/pdf/3.1/qt-whitepaper-us-web.pdf [accessed Juli 1, 2003].

Unwin, A., Wills, G. and Haslett, J. 1990. REGARD --- Graphical Analysis of Regional Data, In *1990 Proceedings of the Section on Statistical Graphics*, American Statistical Association, pp. 36-41, Alexandria, VA.

Westphal, C. and Blaxton, T. 1998: *Data Mining Solutions - Methods and Tools for Solving Real-World Problems*, 8 ed., John Wiley & Sons, Inc, New York.

Wijk, J.J. van and E. van Selow. 1999. Cluster and Calendar-based Visualization of Time Series Data. In: G. Wills, D. Keim (eds.), *Proc. IEEE Symposium on Information Visualization (InfoVis'99)*, IEEE Computer Society, pp 4-9.

AUTHOR BIOGRAPHIES

THOMAS NOCKE is assistant professor at the Institute of Computer Graphics at the University of Rostock. His research interests include visual data mining of climate data and visualization design. You can reach him by e-mail at <nocke@informatik.uni-rostock.de> and his web address is is <wwwicg.informatik.uni-rostock.de/~nocke>.

HEIDRUN SCHUMANN is professor at the Institute of Computer Graphics at the University of Rostock. Her research interests include visualization, visual data mining, rendering and mobile interfaces. You can reach her by e-mail at <schumann@informatik.uni-rostock.de> and her web address is <wwwicg.informatik.uni-rostock.de/~schumann>.

UWE BÖHM is senior scientist at the University Potsdam, Germany. His research focuses on dynamic regional climate modeling and combining uni- and multivariate statistics for model evaluation. His investigations are motivated by the open questions in comprehensively assessing a model's performance and accuracy as a whole. He can be contacted by e-mail at <boehm@pik-potsdam.de>.

MICHAEL FLECHSIG is senior researcher at the Potsdam-Institute for Climate Impact Research in Potsdam, Germany. His research interests include multi-run model simulation and experimentation environments, experiment-specific simulation output analysis and diagnosis for sensitivity and behavioral analyses, Monte-Carlo simulations and scenario runs. You can reach him by e-mail at <flechsig@pik-potsdam.de>.